



# Comparative Genomic Insights Into the Taxonomic Classification, Diversity, and Secondary Metabolic Potentials of *Kitasatospora*, a Genus Closely Related to *Streptomyces*

Yisong Li, Meng Wang, Zhong-Zhi Sun and Bin-Bin Xie\*

State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao, China

## OPEN ACCESS

### Edited by:

Andrés Moya,  
University of Valencia, Spain

### Reviewed by:

Renato Fani,  
University of Florence, Italy  
Liang (Leon) Wang,  
Xuzhou Medical University, China

### \*Correspondence:

Bin-Bin Xie  
xbb@sdu.edu.cn

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 22 March 2021

Accepted: 20 May 2021

Published: 14 June 2021

### Citation:

Li Y, Wang M, Sun Z-Z and  
Xie B-B (2021) Comparative Genomic  
Insights Into the Taxonomic  
Classification, Diversity,  
and Secondary Metabolic Potentials  
of *Kitasatospora*, a Genus Closely  
Related to *Streptomyces*.  
*Front. Microbiol.* 12:683814.  
doi: 10.3389/fmicb.2021.683814

While the genus *Streptomyces* (family *Streptomycetaceae*) has been studied as a model for bacterial secondary metabolism and genetics, its close relatives have been less studied. The genus *Kitasatospora* is the second largest genus in the family *Streptomycetaceae*. However, its taxonomic position within the family remains under debate and the secondary metabolic potential remains largely unclear. Here, we performed systematic comparative genomic and phylogenomic analyses of *Kitasatospora*. Firstly, the three genera within the family *Streptomycetaceae* (*Kitasatospora*, *Streptomyces*, and *Streptacidiphilus*) showed common genomic features, including high G + C contents, high secondary metabolic potentials, and high recombination frequencies. Secondly, phylogenomic and comparative genomic analyses revealed phylogenetic distinctions and genome content differences among these three genera, supporting *Kitasatospora* as a separate genus within the family. Lastly, the pan-genome analysis revealed extensive genetic diversity within the genus *Kitasatospora*, while functional annotation and genome content comparison suggested genomic differentiation among lineages. This study provided new insights into genomic characteristics of the genus *Kitasatospora*, and also uncovered its previously underestimated and complex secondary metabolism.

**Keywords:** *Kitasatospora*, comparative genomics, phylogenomic, inter-genus divergence, diversity, secondary metabolites, recombination

## INTRODUCTION

The genus *Kitasatospora* belongs to the family *Streptomycetaceae*, a high GC multicellular actinobacterial taxon with a complex mycelial life cycle (Goodfellow et al., 2012; Takahashi, 2017). *Kitasatospora* is one of the three major recognized genera within the family, with the other two genera being *Streptomyces*, which prefers neutral to alkaline soils as a natural habitat (Kämpfer, 2012), and *Streptacidiphilus*, representing a group of acidophilic actinobacteria isolated from acidic

soils (Kim et al., 2003). The morphological characteristic of these three genera is rather similar, while *Kitasatospora* is clearly different from the other two in the matter of its cell wall composition as it contains LL- and meso-diaminopimelic acid (DAP), glycine and galactose (Ômura et al., 1982; Takahashi, 2017). Specifically, the taxonomy status of the genus *Kitasatospora* has been considered controversial for many years (Ômura et al., 1982; Wellington et al., 1992; Zhang et al., 1997). Recently, more detailed genetic analyses based on morphology-related genes, including *bldB*, *mbl*, *whiJ*, and *murE*, and also house-keeping genes provided evidence that it should be regarded as a separate genus (Girard et al., 2014; Labeda et al., 2017). To date, more than 33 species in the genus *Kitasatospora* are validly published, many of which have been identified in recent years (Parte, 2018).

Members of *Streptomycetaceae* are well known as rich sources of multifarious secondary metabolites, the majority of which are derived from the genus *Streptomyces* (Hopwood, 2007; Barka et al., 2016). Meanwhile, *Streptomyces* could inhabit various habitats, providing ideal materials for studies of microbial ecology and evolution. Therefore, *Streptomyces* species have received enormous attention in microbial sampling efforts. It has also been shown that *Streptomyces* genomes are recombinogenic, with a high recombination rate exceeding those observed within many other bacterial species (Doroghazi and Buckley, 2010). As the sister genus of *Streptomyces*, *Kitasatospora* has gradually attracted people's attention (Hao et al., 2020; Kim et al., 2020; Yun et al., 2020). So far, at least 50 bioactive compounds have been discovered from *Kitasatospora* strains (Takahashi, 2017), such as setamycin, kitasetaline, and satorporin (Ômura et al., 1981; Aroonsri et al., 2012; Arens et al., 2013), exhibiting a wide range of bioactivities (e.g., antifungal, herbicidal, antitumor). Nevertheless, little is known about the potential differences in the secondary metabolism and genetics between genera.

This study was aimed (1) to investigate the phylogenetic and taxonomic position of *Kitasatospora*, and (2) to explore the secondary metabolic potentials of *Kitasatospora*. To address the former issue, phylogenomic analyses were performed and genome content were compared between *Kitasatospora* and the closely related genera. To address the later, secondary metabolite biosynthetic gene clusters were annotated and compared in *Kitasatospora* genomes. This study represented the first comprehensive comparative genomic study of the genus *Kitasatospora*.

## MATERIALS AND METHODS

### Genome Data Set

All genomes of the members of the family *Streptomycetaceae* (April 2020) were downloaded from the NCBI genomes FTP site<sup>1</sup>. Genomic qualities of those genomes were assessed by using CheckM (Parks et al., 2015), and only genomes that were at least 95% complete and had no more than 5% contamination were screened to the further study. Taxonomy assignment of these

genomes was performed using the Genome Taxonomy Database (GTDB) toolkit (Chaumeil et al., 2019). To ensure that each genus contained enough genomes to allow statistically robust analysis, genomes were dereplicated using dRep (Olm et al., 2017), with 99.5% average nucleotide identity (ANI) for *Kitasatospora* and 95% for *Streptomyces* and *Streptacidiphilus*. Through these, a total of 45, 575, and 12 high-quality and non-redundant genomes of strains belonging to *Kitasatospora*, *Streptomyces*, and *Streptacidiphilus* were retained. The information of isolation source was obtained from the NCBI BioSample database and from the literature. Similarity searches for the 16S rRNA gene sequences were performed by the EzBioCloud server (Yoon et al., 2017). Whole genome ANI for each pair of genomes was computed by FastANI (Jain et al., 2018).

### Genome Annotation and Determination of Homologous Gene Families

To avoid bias due to differences in gene calling and annotation, all assemblies were reannotated using Prokka (Seemann, 2014). To improve the accuracy of genome annotation, function annotation and classification of proteins were performed by sequence comparison using DIAMOND BLASTP (*E*-value 1e-05, coverage 0.5, and identity 50%) (Buchfink et al., 2015) against the recently updated Clusters of Orthologous Group (COG) database (Galperin et al., 2020). If a gene was assigned to more than one COG category, each COG category was calculated separately. PFAM domains of proteins were further identified using PfamScan (El-Gebali et al., 2019) with default parameters. Insertion sequences were identified by BLASTP against the ISFinder database (*E*-value 1e-05) (Siguier et al., 2006). Via the use of GET\_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013) with OrthoMCL (Li et al., 2003) clustering algorithm, homologous gene families were calculated, and cloud, shell, and (soft-) core pangenome components were also derived. Pan-genome statistics were visualized using PanGP (Zhao et al., 2014).

### Phylogenomics and Population Structure Analyses

To build the phylogenetic tree of the family *Streptomycetaceae*, we used the GToTree (Lee, 2019) package with default parameters and the "Actinobacteria" HMM-set. Through this, a concatenated protein alignment from the 138 marker genes was constructed using Muscle (Edgar, 2004) and subsequently trimmed using TrimAl (Capella-Gutiérrez et al., 2009). The approximately maximum-likelihood phylogenetic tree was built with FastTree (Price et al., 2010). We also constructed a marker-based phylogenetic tree of the genus *Kitasatospora* by using the GET\_PHYLOMARKERS pipelines (Vinuesa et al., 2018) run in default mode, identifying high-quality marker genes for robust phylogenomic analyses based on nucleotide sequences of the genes from the resulting single-copy core genome. Within the pipelines, the final phylogenetic tree was generated by ML algorithms with the best model of evolution GTR + F + ASC + R4 in the IQ-TREE software (Nguyen et al., 2015) based on the concatenated gene sequences with

<sup>1</sup><https://www.ncbi.nlm.nih.gov/genomes>

1,000 bootstrap replicates. The tree topology was visualized using the Interactive Tree of Life (Letunic and Bork, 2016). In addition, we used R function `pvclust` (Suzuki and Shimodaira, 2006) to perform a hierarchical cluster analysis, based on an absence/presence (0/1) matrix of dispensable genes according to GET\_HOMOLOGUES results. K-medoids clustering was performed by using the partitioning around medoids (PAM) algorithm from the R package `fpc`<sup>2</sup>. Population structure was assessed using hierarchical Bayesian analysis of population structure (BAPS) (Corander et al., 2004).

## Genome Mining for Secondary Metabolite Biosynthetic Gene Clusters

Prediction of secondary metabolite biosynthetic gene clusters (smBGCs) was performed using the antiSMASH 5.0 software (Blin et al., 2019), with annotations based on the MIBiG repository (Medema et al., 2015). The final BGC set was analyzed using BiG-SCAPE (Navarro-Muñoz et al., 2020) under `-hybrid` and `-mode glocal` settings at distance cutoffs of 0.3, 0.5, and 0.7. The networks were visualized with Cytoscape (Shannon et al., 2003). We applied COUNT (Csürös, 2010) under the Wagner parsimony algorithm for ancestral reconstruction of secondary metabolite biosynthetic gene cluster families (smBGCs), and for detecting the gain and loss events that might have occurred during the evolutionary history of the genus *Kitasatospora*.

## Estimation of Genetic Recombination

SplitsTree (Huson and Bryant, 2006) was used to construct the network with the NeighborNet algorithm, and also to calculate the pairwise homoplasy index (Bruen et al., 2006). Ancestral and recent recombination events happening in each single-copy core gene were identified using fastGEAR (Mostowy et al., 2017) with default parameters. Recombination parameters, including the fraction of sample diversity derived from recombination ( $c$ ) and the relative rate of recombination to mutation ( $\gamma/\mu$ ) were calculated by `mcorr` (Lin and Kussell, 2019) with default parameters, based on the single-copy core genes. Fifty subsets were constructed, each containing 50 different species (pairwise ANI < 95%; 20 for genus *Kitasatospora*, 20 for genus *Streptomyces*, and 10 for genus *Streptacidiphilus*) randomly selected from the genome dataset.

## Statistical Analysis

Variation in the functional profiles was performed using non-metric multidimensional scaling (NMDS) analysis on Bray-Curtis dissimilarity matrices with the “ordinate” function in the phyloseq R package (McMurdie and Holmes, 2013). The significance of the differences in functional profiles was tested by permutational multivariate analyses of variance (PERMANOVA) using the “adonis” function in the vegan R package<sup>3</sup> with 999 random permutations. SIMPER analysis was employed with the “simpler” function of the vegan R package to determine the most differentiating COG entries among different genome groups.

<sup>2</sup><https://cran.r-project.org/web/packages/fpc/>

<sup>3</sup><https://github.com/vegandevs/vegan>

Statistical analyses, including Wilcoxon test, Kruskal-Wallis test, and Student’s *t*-test, were performed in R.

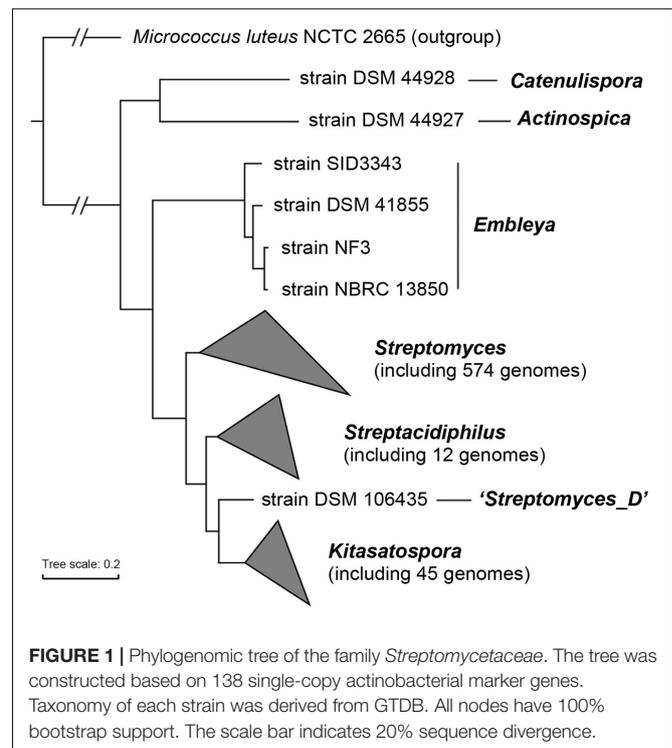
## RESULTS

### Phylogenetic Position of *Kitasatospora*

Our dataset contained genomes for 45 *Kitasatospora* strains, including five complete and 40 draft genome sequences, and the characteristic genomic features are summarized in **Supplementary Table 1**. Using 138 single-copy actinobacterial marker genes, we examined the phylogenetic relatedness of the entire family *Streptomycetaceae* (Figure 1). The result showed a good correlation with those observed based on 16S rRNA gene and multi-locus sequences (Labeda et al., 2017; Takahashi, 2017). Generally, strains were clustered based on their source of genus, supporting the current taxonomic status of *Kitasatospora* as a separate genus. The only exception was strain DSM 106435, which has been proposed as the type strain of *Streptacidiphilus bronchialis* (Nouioui et al., 2019) and has been classified as “*Streptomyces\_D*” in GTDB. However, phylogenomic analysis indicated that this strain was more closely related to *Kitasatospora*.

### Inter-Genus Comparison

The genome size of *Kitasatospora* ranged from 6.3 (strain NRRL ISP-5024) to 12.3 Mbp (strain CB01881), with a mean size of  $8.7 \pm 1.0$  Mbp, similar to those of the other two genera (Supplementary Figure 1A). Notably, the G + C content varied between 71.2 and 74.5% with an average of  $73.1 \pm 0.8\%$ ,



which was significantly larger than those of *Streptomyces* and *Streptacidiphilus* ( $71.7 \pm 0.9\%$  and  $71.8 \pm 0.7\%$ , respectively; Kruskal-Wallis test,  $p < 0.001$ ; **Supplementary Figure 1B**). Calculation of ANI revealed higher intra-genus ANI values of *Kitasatospora* ( $>85.7\%$ ) than inter-genus values ( $<81.7\%$  for *Kitasatospora-Streptomyces* and  $<80.9\%$  for *Kitasatospora-Streptacidiphilus*), suggesting boundaries between genera. It was noted strain DSM 106435 showed ANI values of  $\sim 81.1\%$  (coverage,  $\sim 43.5\%$ ) with the *Kitasatospora* clade and of  $\sim 80.2\%$  (coverage,  $\sim 46.2\%$ ) with the *Streptacidiphilus* clade, all of which were close to the above inter-genus values. This result, together with the unique phylogenetic position of strain DSM 106435, suggested that strain DSM 106435 may represent a new genus.

Clusters of Orthologous Group annotations of 45 *Kitasatospora* and 587 non-*Kitasatospora* representative genomes were used as input to assess the correlation and divergence of functional traits between genera. Based on the proportions of each COG category, no significant differences were detected between the three genera (PERMANOVA test, the significance level  $p = 0.05$ ; **Figure 2A**). However, based on the number of proteins annotated to each COG entries, the three genera showed different functional profiles (PERMANOVA test,  $p = 0.001$ ; **Figure 2B**).

A total of 31 COGs were detected that significantly contributed most to the dissimilarity between genera (SIMPER analysis,  $>0.05\%$  contribution,  $p < 0.01$ ; **Figure 2C** and **Supplementary Table 2**). Particularly, about half (14/31) functional attributes were related to transcription (COG K), lipid transport and metabolism (COG I) and posttranslational modification, protein turnover, or chaperones (COG O). For example, five transcription/response related gene families (COG2771, COG0640, COG1846, COG1802, and COG2197) made great contribution to the separation of *Kitasatospora* and *Streptomyces*: *Kitasatospora* contains more genes of transcriptional regulator families CsgD, ArsR, MarR and NarL/Fix], while *Streptomyces* contains more GntR, suggesting potential differences in the regulatory mechanisms. For lipid metabolism, two differentiating COG entries (COG0304 and COG0236), implicated in type II fatty acid biosynthesis, were more abundant in *Kitasatospora*, whereas COG1960, COG1024 (both involved in the carnitine metabolism) and COG0183 (participating in the Acetyl-CoA acetyltransferase activity) were more abundant in the other two. There were also four COGs related to posttranslational modification, protein turnover, or chaperone functions, including COG3227 (Zn-dependent metalloprotease), COG1670 (protein N-acetyltransferase), COG4934 (serine protease), and COG1404 (serine protease). Among them, COG4934 has been reported to pertain to acidity linked genes in archaea (Gonzalez and Zimmer, 2008). Consistently, as the most acidophilic genus in *Streptomycetaceae* (Goodfellow et al., 2012), *Streptacidiphilus* has the largest number of copies (10.6, averagely), compared with only 1.6 and 4.7 copies in *Streptomyces* and *Kitasatospora*, respectively. In addition, genes of COG1545, which encoded an uncharacterized OB-fold protein, only existed in 16 (35.56%) *Kitasatospora* strains with only one copy (pairwise amino acid sequence identity  $> 52.10\%$ ), but were presented in all strains of *Streptacidiphilus* and 73.39%

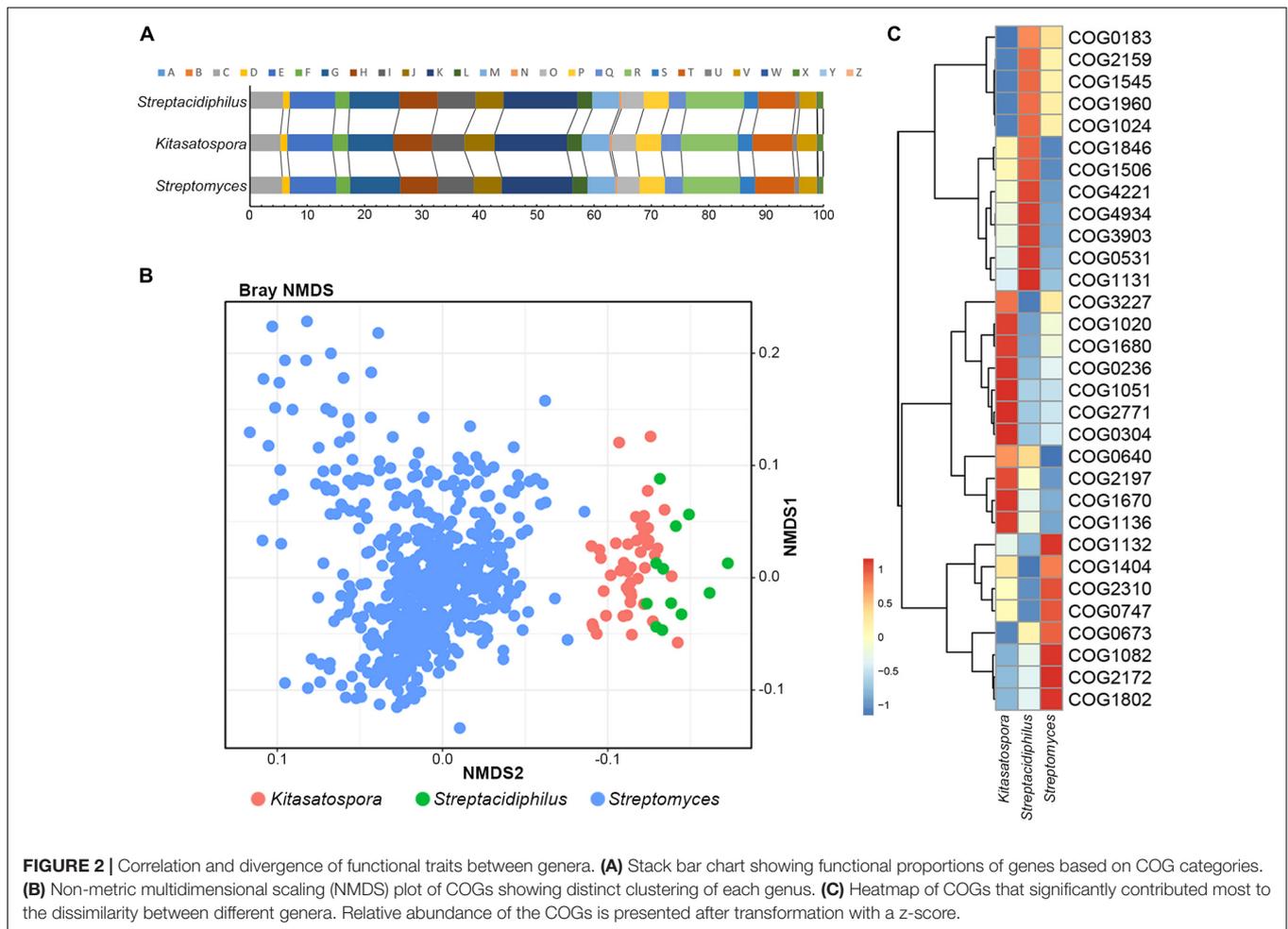
of strains of *Streptomyces*, with an average of 5.58 (median, 5) and 4.87 (median, 5) copies, respectively, and shared a low identity (amino acid sequence identity  $< 45.20\%$ ) with the former.

## Intra-Genus Diversity

The number of coding sequences per genome of *Kitasatospora* ranged between 5,482 and 9,863 (mean,  $7,568 \pm 748$ ). Based on the gene content table obtained by GET\_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013), the genus *Kitasatospora* had a pan-genome of 55,080 gene families (**Figures 3A,B**). Among these, 1,476 families (2.7% of total) were presented in all strains (core genome), and 2,563 families (4.7%, including the core genome) were conserved in more than 93% of all strains analyzed here (soft-core genome), which represented 26.3 to 47.1% of the gene content of each strain (35.7%, averagely). The cloud genes (rare genes presented only in one or two strains) contained 70.5% of the pan-genome (38,847 families) and accounted for 13.5% for each strain, averagely. The remaining 24.8% were shell genes (13,670), making up about half the genome content for each strain, averagely. Thus, the flexible content (cloud plus shell genes) formed more than 95% of the pan-genome. The pan-genome fitting curve did not reach a plateau as more genomes were sampled ( $\alpha = 0.61$ , estimated by power-law regression), suggesting that the pan-genome of *Kitasatospora* was open (**Figure 3B**). Correspondingly, the pan-genome still increased with more than 700 new genes after addition of a 45th genome. This pan-genome structure was comparable to that of *Streptomyces* (**Figures 3C,D**). All these results indicated that the gene content of the genus was highly diverged.

To further evaluate the intra-genus differentiation of the genus, phylogenomics reconstruction was performed based on 456 high-quality phylogenomic marker (core) genes. As shown in **Figure 4**, the phylogenomic tree was well-supported, showing three well-defined genetic clades: Clades I, II, and III. A hierarchical clustering tree based on the content of dispensable genes showed three clusters generally corresponding to the three clades in the core gene tree (**Supplementary Figure 2**), indicating the existence of inter-clade gene content difference within the genus. In the core gene tree, 15 strains formed a monophyletic lineage (**Figure 4**, purple, named as Lineage IV) that diverged deeply from the other strains (**Figure 4**, cyan, Lineage III, paraphyletic) in Clade III. Similarly, these 15 strains were clustered into a single branch (**Supplementary Figure 2**, purple) that was different from the main branch of other strains (**Supplementary Figure 2**, cyan) from Clade III of the core gene tree. This result was generally consistent with the k-medoids clustering and population structure analyses, both of which showed differences within the Clade III (**Supplementary Table 1**). Therefore, in addition to the inter-clade differentiation, differentiation between Lineages IV and III (both within Clade III) has merged. The intra-genus difference was further investigated based on the four lineages: Lineages I (corresponding to Clade I), II (corresponding to Clade II), III, and IV.

We next sought to detect the inter-lineage differences. Genome sizes of the lineages were largely same (**Figure 5A**). The G + C content of Lineage IV was higher than those of



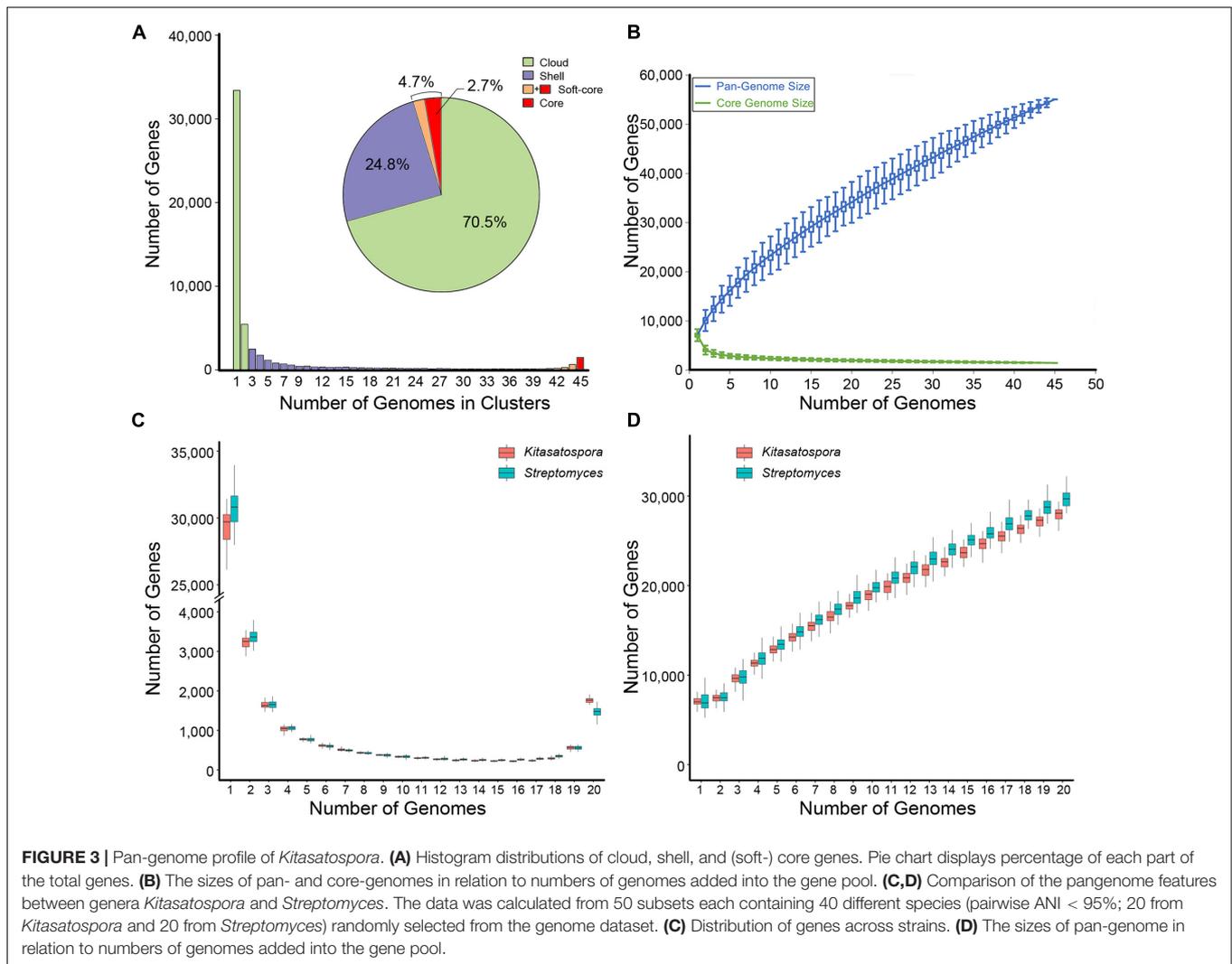
others (Figure 5B). Functional profiles were also significantly different between the four lineages (PERMANOVA,  $p = 0.001$ ; Figure 5C). A total of 35 COGs were detected which greatly promoted the separation between lineages (Figure 5D), ten of which also showed a differential distribution among lineages within *Kitasatospora* (Supplementary Table 3). For example, COG4934, the acidity linked genes mentioned before, showed similar abundance in Lineage I to that in *Streptacidiphilus*. This gene was presented with on average 10.2 copies in Lineage I, while less than 5.2 in Lineages II to IV, suggesting that Lineage I might be more suitable for acidic environments than other lineages. Similarly, a predicted ATPase-coding gene (COG3903) was also more abundant in Lineage I and *Streptacidiphilus* than other lineages.

## Great Potential and Diversity of Secondary Metabolism in *Kitasatospora*

To assess the genomic repertoire for secondary metabolite biosynthetic gene clusters (smBGCs), we performed a computational genome mining analysis using the program antiSMASH (Blin et al., 2019). A total of 1,535 smBGCs were predicted, with 22 – 50 (mean,  $34.1 \pm 8.2$ ) smBGCs per genome,

and the total length of these gene clusters occupied 8.7 – 24.7% (mean,  $15.6 \pm 3.9\%$ ) of the whole genome (Supplementary Figure 3). The number and percentage of length in *Kitasatospora* were significantly larger than those in *Streptomyces* ( $30.9 \pm 7.9$ ;  $13.2 \pm 4.0\%$ ; Wilcoxon test,  $p = 0.027$  and  $p = 0.00014$ , respectively) and *Streptacidiphilus* ( $27.8 \pm 6.8$ ;  $10.6 \pm 3.2\%$ ;  $p = 0.068$  and  $p = 5.816e-05$ , respectively). Difference was also observed between the *Kitasatospora* lineages (Kruskal-Wallis test,  $p = 0.0025$ ): Lineage II contains on average more smBGCs than Lineages III and IV (Wilcoxon test,  $p < 0.05$ ; Supplementary Figure 4). These 1,535 *Kitasatospora* smBGCs could be grouped into 8 major biosynthetic classes (Supplementary Figure 4), with an average of 19.6% RiPPs, 17.2% NRPS, 14.7% terpene, 7.3% PKS-NRP\_Hybrids, 6.5% “PKSother”, 4.7% type I PKS, 0.25% saccharide, and 29.6% others (non-classifiable) in each strain. Compared with *Streptomyces*, *Kitasatospora* are significantly richer in classes RiPPs, NRPS, and “PKSother”, but poorer in “Others” (Wilcoxon test,  $p < 0.05$ ).

In order to gain insights into the diversity of smBGCs, similarity networks were constructed using the BiG-SCAPE algorithm (Navarro-Muñoz et al., 2020), which defined “secondary metabolite biosynthetic gene cluster families” (smBGCfs) based on their Pfam similarities. A total of 896, 245,

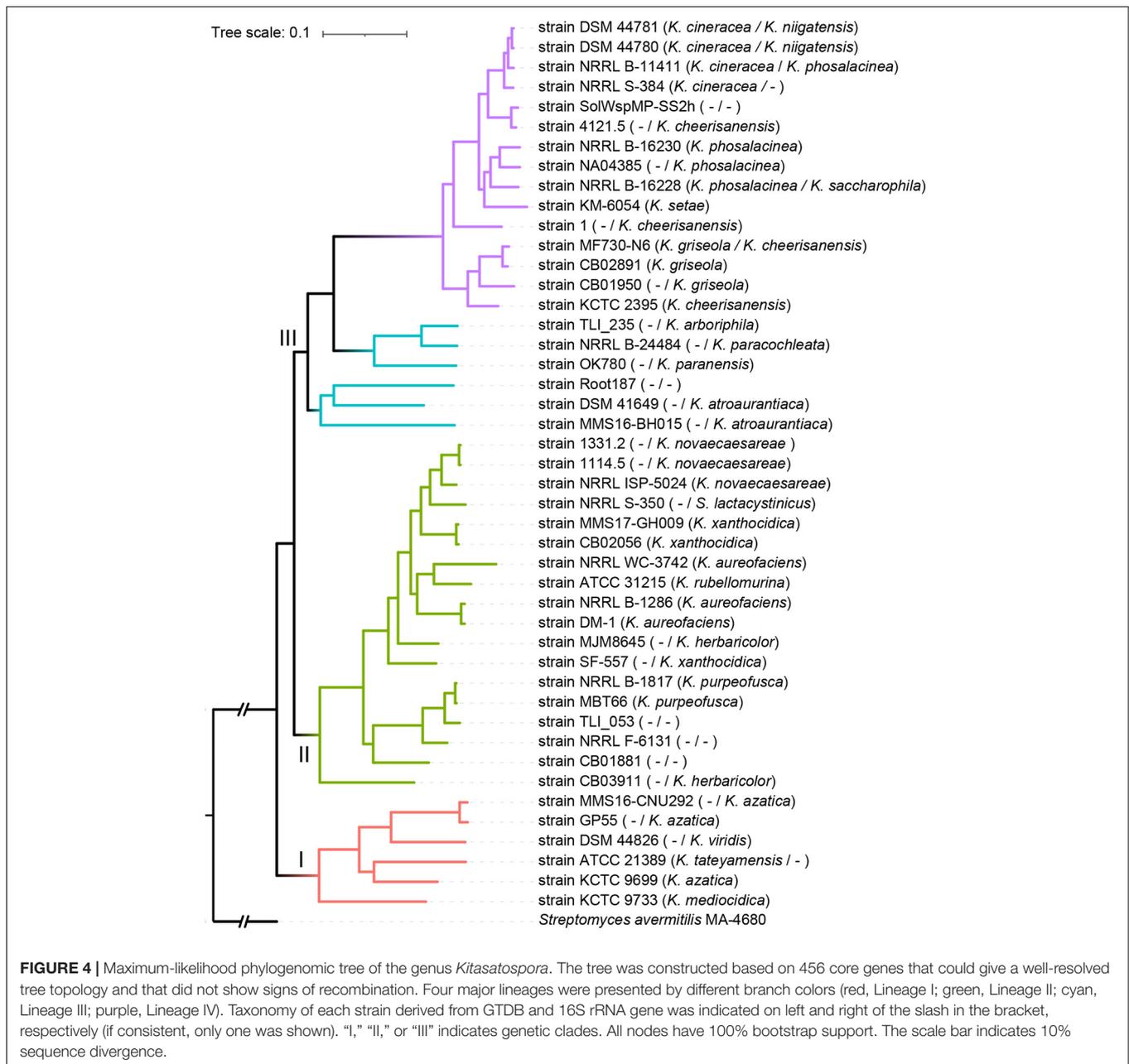


and 9,215 smBGCFs were detected ( $c = 0.3$ ) for *Kitasatospora*, *Streptacidiphilus* and *Streptomyces*, respectively. Comparison of the presence and absence of these smBGCFs revealed that the three genera had highly divergent secondary metabolic profiles (PERMANOVA,  $p = 0.001$ ). For the genus *Kitasatospora*, up to 92.9% of the total smBGCFs were genus-specific, and only 2.3% shown homology to MIBiG entries. Only 2.2% ( $n = 20$ ) were conserved in more than 20% of strains, and more than three-quarters ( $n = 692$ , 77.2%) were singletons, suggesting a broad secondary metabolic diversity within the genus. Even by raising the default similarity score cutoff from  $c = 0.3$  to  $c = 0.7$ , the proportion of conserved smBGCFs did not change a lot (5.9%), and still more than half (51.4%) were singletons.

To further assess the smBGCFs diversity in *Kitasatospora*, we chose to use a relatively stricter raw distance cutoff of 0.3. We firstly investigated the gain/loss events of smBGCFs that might have occurred during the evolutionary history of the genus *Kitasatospora* (Supplementary Figure 5). The family numbers showed a noticeable upward trend, with 1,043 gains while only 32 losses. No smBGCF was inferred at the last common ancestor

of the genus, and no more than three were inferred at the nodes of the three major clades. We also noticed that 81.40% of total gain events and 62.50% of total loss events occurred at terminal nodes. In addition, within each genome, between 14.29 and 42.86% of the smBGCFs (mean, 26.92%) comprised one or more genes that were annotated as mobile genetic elements (MGEs) such as insertion sequences, prophages or transposases. These results suggested that the smBGCFs of *Kitasatospora* have experienced remarkable expansion, perhaps mediated by HGT, in the evolutionary history.

The network of non-singleton smBGCFs was also built (Figure 6 and Supplementary Figure 6). A total of 178 networks were formed. Two networks, networks 1 and 2, contained smBGCFs from all the 45 *Kitasatospora* strains. Network 1 included two types of clusters: a hopene biosynthetic gene cluster, which was found in all *Kitasatospora* genomes, and a type III PKS gene cluster, which was mainly distributed in Lineage IV and included three conserved biosynthetic genes from the alkylresorcinol biosynthetic gene cluster. Network 2 summarized pathways that were likely responsible for the formation of an

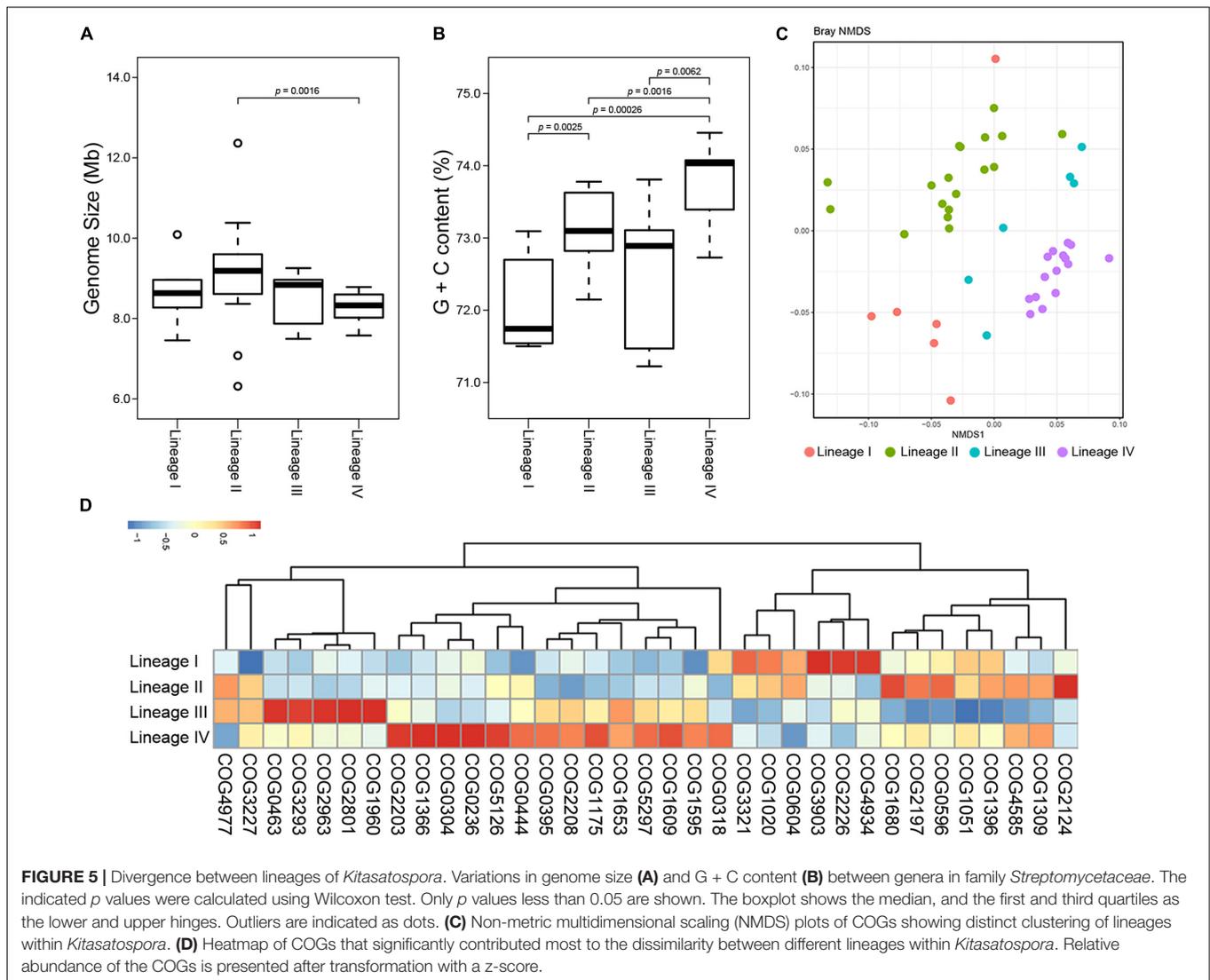


uncharted siderophore. The main genetic structures of these three clusters above were highly homologous to those in *Streptomyces*. Only a small number of networks contained members that showed similarities to reported clusters. For example, smBGCs from Network 7 were predicted to encode unknown linear azole-containing peptides (LAPs), which displayed similarities with those in *Verrucosipora maris* and *Herbidospora* spp.; Networks 10, 16, and 78 contained predicted spore pigment polyketide gene clusters; and Network 47, contained a well-studied desferrioxamine B gene cluster, which was very common in *Streptomyces*. Other well-annotated smBGCs also included members encoding streptobactin/griseobactin (Network 23), bafilomycin (Network 14), microansamycin (Network 30),

and endophenazine (Network 38). Furthermore, most of the abundant *Kitasatospora*-specific networks were lineage-specific (Figure 6, networks with nodes in the same color). These results revealed a huge diversity and potential of secondary metabolite production that may be harnessed through dedicated mining of *Kitasatospora*.

## High Frequency of Recombination in *Kitasatospora* Genomes

Recombination is known to contribute to the evolution of streptomycetes, which has been reported to frequently recombine within and between species (Doroghazi and Buckley, 2010;

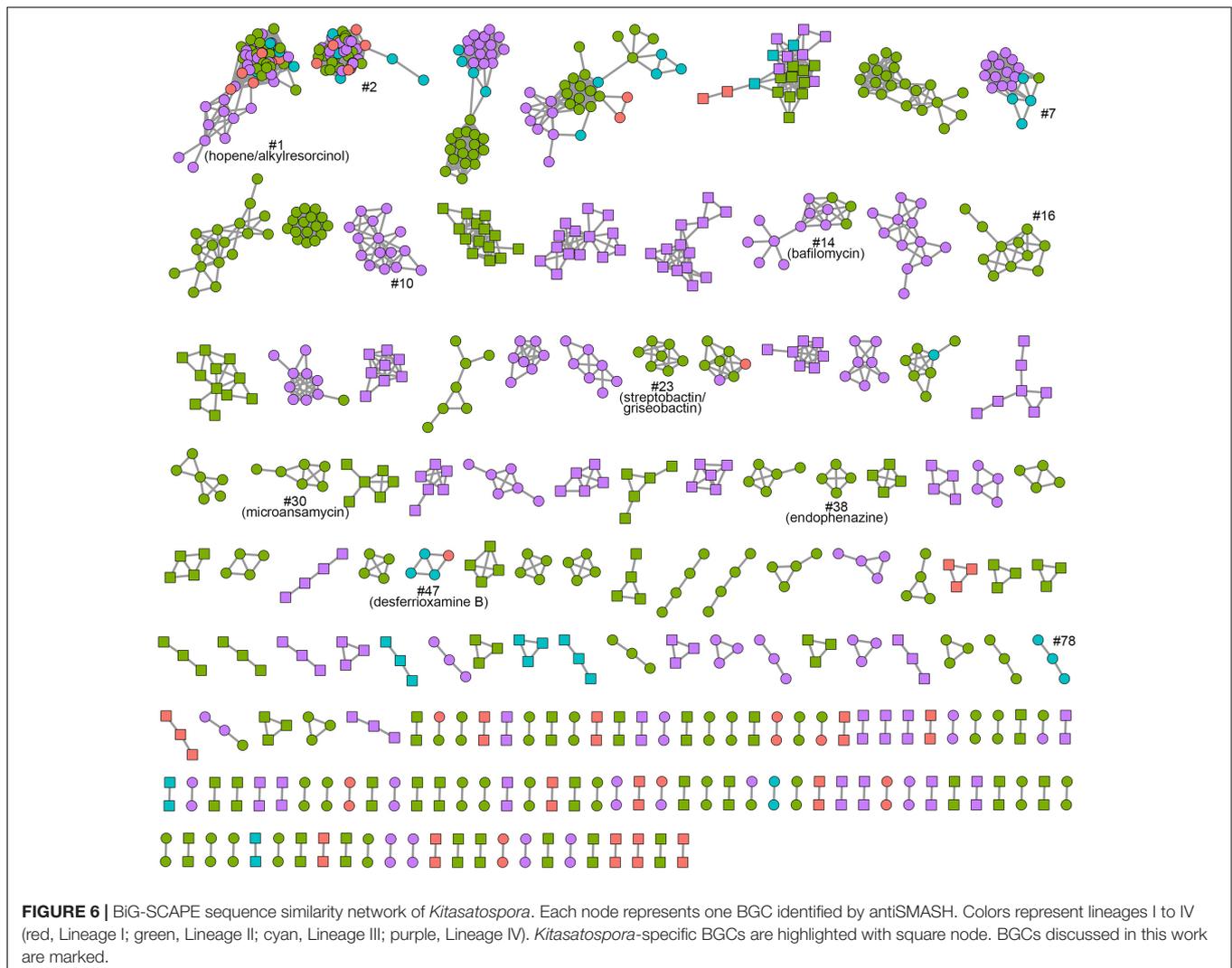


Cheng et al., 2016). Here, we sought to determine the frequency and extent of recombination in *Kitasatospora*. Firstly, calculation of the pairwise homoplasy index (PHI) (Bruen et al., 2006) revealed evidence for significant recombination in the core genome (1,274 single-copy core genes; *p*-value = 0.0). This could be visualized by NeighborNet network (Figure 7A), which showed a reticulate structure due to non-vertical inheritance in phylogeny (Huson and Bryant, 2006). This observation was further supported by the fastGEAR analysis (Mostowy et al., 2017), which showed that more than 98% of the core genes have been affected by recombination in the evolutionary history. In total, 1,243 (97.6%) genes were involved in recent recombination and 518 (40.7%) genes in ancestral recombination (Figure 7B). Specially, genes *pulA* (encoding a pullulanase, EC 3.2.1.41), *lpqB* (encoding a putative lipoprotein), and *mrcB* (encoding a penicillin-binding protein) have undergone 119, 98, and 74 recent recombinations, respectively. FAA1 (encoding a long-chain fatty acid-CoA ligase) and *aprE* (encoding an alkaline serine protease) have undergone 13 and 11 ancestral recombinations.

We also sought to quantify the impact of recombination on the genetic diversity of *Kitasatospora* (Figures 7C,D). The recombination coverage (*c*), the fraction of the genome whose diversity was derived from recombination events since the last common ancestor of the population, was estimated to be  $0.90 \pm 0.003$ . The relative rate of recombination to mutation ( $\gamma/\mu$ ) was estimated to be  $6.7 \pm 0.90$ . Interestingly, there were significant differences in these two parameters between the three genera (*p* < 0.001, paired two-tailed Student's *t*-test): *Kitasatospora* has the lowest *c* value, but the highest  $\gamma/\mu$  ratio.

## DISCUSSION

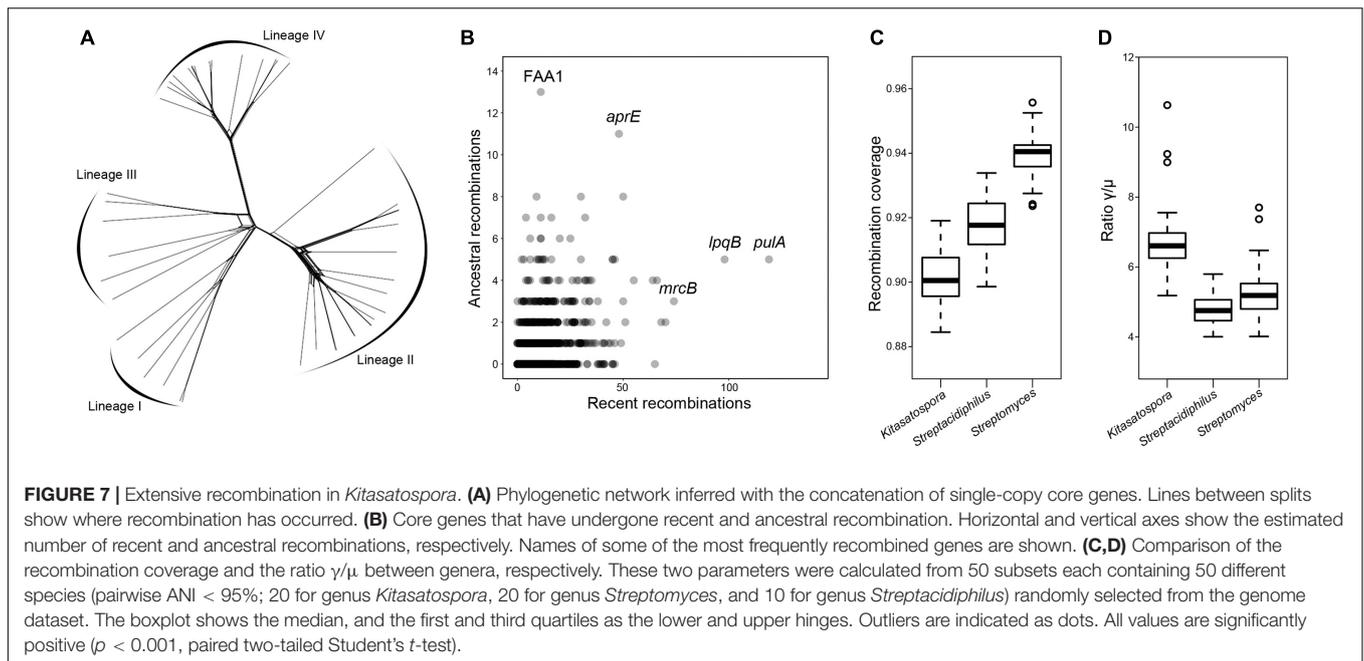
While *Streptomyces* is the model of bacterial physiology and an important resource of antibiotic discovery, its close relatives are rarely studied. To the authors' knowledge, this study represented the first systematic comparative genomic study of the *Kitasatospora* genus, a close relative of *Streptomyces*.



Our phylogenomic analysis confirmed the taxonomic status of *Kitasatospora* as a separate genus distinct from *Streptomyces*. This result is consistent with the studies based on multi-locus sequence analysis (Labeda et al., 2017). Our phylogenomic analysis also revealed discrepancies between the classification and phylogenetic positions for a few strains, suggesting potential classification errors (Figures 1, 4). For example, strain DSM 106435, the type strain of *Streptacidiphilus bronchialis* (Nouioui et al., 2019), may represent a new genus (Figure 1). Furthermore, our analyses also revealed differences between the GTDB classification and the 16S rRNA gene-based taxonomy (Ezbiocloud). For example, strains DSM 44781, DSM 44780, NRRL B-11411, and NRRL S-384 were all assigned to *K. cineracea* with inner ANI > 95.8% by GTDB, but they were identified into three different species by Ezbiocloud based on 16S rRNA gene, although the gene identity > 98.5%. Strains MF730-N6 and CB02891 were assigned as *K. cheerisanensis* and *K. griseola* based on 16S rRNA gene, respectively. However, the 16S rRNA gene identity between these two strains was > 99.8% and the pairwise ANI value was > 97.3%, suggesting these two strains

should belong to the same species. Generally, phylogenomics has surpassed 16S rRNA gene-based approaches in terms of elucidating phylogenetic relationships and facilitated accurate (re)classification of microbes (Zhang et al., 2016; Hinger et al., 2020).

Our comparative genomic analyses also revealed genome content differences between genera. Interestingly, the abundance of genes of COG4934 (serine protease, subtilase family) showed different abundance in different genera: on average 10.6 copies in *Streptacidiphilus*, 4.7 copies in *Kitasatospora*, and 1.6 copies in *Streptomyces*. COG4934 was among the top ranked COGs that were correlated with acidophilic bacteria and archaea (Gonzalez and Zimmer, 2008), and has been considered as the acidity linked gene, although the molecular mechanism has not been determined. The genus *Streptomyces* is regarded to prefer neutral to alkaline environmental pH (optimal pH range: 6.5–8.0), although they commonly occur at remarkably variable pH conditions, and the genus *Streptacidiphilus* is more acidophilic with optimum growth at pH 4.5–5.5, while the preferred pH for genus *Kitasatospora* has not been determined



(Goodfellow et al., 2012). Based on these, we speculated that the optimal growth pH of the genus *Kitasatospora* is between those of genera *Streptomyces* and *Streptacidiphilus*. In addition, this COG showed similar abundance in *Kitasatospora* Lineage I to in *Streptacidiphilus*. Therefore, it is possible that, similar to *Streptacidiphilus*, *Kitasatospora* Lineage I prefers acidic environments. Since *Kitasatospora* is closely related to *Streptacidiphilus* and Lineage I is the earliest diverged lineage in *Kitasatospora*, it is possible that the last common ancestor of *Kitasatospora* and *Streptacidiphilus* had abundant COG4934 and a few *Kitasatospora* lineages lost a few copies of COG4934 after their divergence from Lineage I. In addition, we have also detected a series of COGs that perhaps associated with phenotypic traits (type II fatty acid biosynthesis, carnitine metabolism, etc). We also noticed that genes of COG1545 that were largely absent in most *Kitasatospora* strains but were quite abundant in the other two genera. It has been reported that proteins of COG1545 were involved in synthesis of 2,4-Diacetylphloroglucinol, a low-molecular weight polyketide antibiotic produced by *Pseudomonas brassicacearum* strains (Mandryk-Litvinkovich et al., 2017), suggesting another obvious difference between genera. Nevertheless, more experimental evidences, both morphologically (e.g., morphological characteristics revealed by transmission electron microscopy) and biochemically, were needed to verified these differences, especially between *Kitasatospora* and *Streptomyces*.

Though there were inter-genus genome content differences, *Kitasatospora* genomes possessed high numbers of smBGCs, which were higher than those from *Streptomyces* and *Streptacidiphilus*. Since the cluster number may be affected by broken or merged clusters due to low genome assembly quality, we checked the potential correlations between the number of

smBGCs and the number of contigs per genome. It was shown that the smBGC number was not correlated with the contig number ( $R^2 = 0.078$ ,  $p = 0.035$ ), but rather positively correlated with the genome size ( $R^2 = 0.42$ ,  $p = 8.0e-07$ ). Similarly, the total smBGC length and ratio of smBGC length were also bigger in *Kitasatospora* than the other two genera. A noticeable upward trend of these smBGCs were found during the evolutionary history, and most of the gene gain events occurred at terminal nodes, suggesting HGT events that happened recently may increase the diversity of secondary metabolism in *Kitasatospora*. Furthermore, it was also noted that most of smBGCs found in *Kitasatospora* were different from *Streptomyces* and from those reported, strongly suggesting *Kitasatospora* is a valuable and new source for antibiotic discovery in the future.

*Streptomyces* has long been considered as an interesting model for exploring patterns of microbial differentiation in terms of genetic and ecological diversity. Until now, progresses have been made in studies of *Streptomyces* evolution, revealing a high rate of horizontal gene transfer and also a widespread homologous recombination (Andam et al., 2016; McDonald and Currie, 2017; Tidjani et al., 2019). However, little is known about *Kitasatospora*. In this study, the pan-genome analyses revealed a massive gene flux in *Kitasatospora* comparable to that of *Streptomyces*. Estimation of recombination events revealed high level of recombination in *Kitasatospora* as well as in the two related genera, and the recombination coverage (c) and rate of recombination to mutation ( $\gamma/\mu$ ) are among the highest values for bacteria reported so far (Lin and Kussell, 2019; Park and Andam, 2020; Li et al., 2021).

Collectively, this study provides comprehensive comparative genomic insights into the assessment of the phylogenomic relationship and diversity of members of the genus *Kitasatospora*, and also provides evidence that supported *Kitasatospora* as a

separate genus, with genomic features different from other genera in the family. This study also suggests that *Kitasatospora* is a valuable resource for the natural product discovery, which should be further experimentally confirmed in the future.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. All data analyzed during this study are available through NCBI GenBank database, and are accessible through the accession numbers listed in **Supplementary Table 1**.

## AUTHOR CONTRIBUTIONS

YL and B-BX contributed to conception and design of the study. YL performed the main bioinformatic analyses and wrote the draft manuscript. MW, Z-ZS, and B-BX interpreted and discussed the results and revised the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This study was supported by the National Natural Science Foundation of China (31770412), the Young Scholars Program of Shandong University (2016WLJH41), and Youth

Interdisciplinary Science and Innovative Research Groups of Shandong University (2020QNQT006).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.683814/full#supplementary-material>

**Supplementary Figure 1** | Variations in genome size (A) and G + C content (B) between genera in family *Streptomycetaceae*.

**Supplementary Figure 2** | Hierarchical cluster analysis based on the presence or absence of dispensable genes.

**Supplementary Figure 3** | Comparison of the total number (A), the total size (B), and percentage of size (C) of smBGCs between genera.

**Supplementary Figure 4** | Distribution of biosynthetic gene clusters among strains of *Kitasatospora*.

**Supplementary Figure 5** | Ancestral reconstruction of smBGCs in *Kitasatospora* using COUNT software.

**Supplementary Figure 6** | Genetic organizations of the BGCs discussed in the work.

**Supplementary Table 1** | Genomic features of *Kitasatospora* strains used in this study.

**Supplementary Table 2** | COGs contributing most significantly ( $>0.05\%$ ,  $p < 0.01$ ) to differences across genera.

**Supplementary Table 3** | COGs contributing most significantly ( $>0.05\%$ ,  $p < 0.01$ ) to differences across lineages within *Kitasatospora*.

## REFERENCES

- Andam, C. P., Doroghazi, J. R., Campbell, A. N., Kelly, P. J., Choudoir, M. J., and Buckley, D. H. (2016). A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *MBio* 7, e02200–e02215. doi: 10.1128/mBio.02200-15
- Arens, J. C., Berrue, F., Pearson, J. K., and Kerr, R. G. (2013). Isolation and structure elucidation of satorsporin A and B: new polyketides from *Kitasatospora griseola*. *Org. Lett.* 15, 3864–3867. doi: 10.1021/ol401598f
- Aroonsri, A., Kitani, S., Hashimoto, J., Kosone, I., Izumikawa, M., Komatsu, M., et al. (2012). Pleiotropic control of secondary metabolism and morphological development by KsbC, a butyrolactone autoregulator receptor homologue in *Kitasatospora setae*. *Appl. Environ. Microbiol.* 78, 8015–8024. doi: 10.1128/AEM.02355-12
- Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H. P., et al. (2016). Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* 80, 1–43. doi: 10.1128/MMBR.00019-15
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: 10.1093/nar/gkz310
- Bruen, T. C., Philippe, H., and Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172, 2665–2681. doi: 10.1534/genetics.105.048975
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019:848. doi: 10.1093/bioinformatics/btz848
- Cheng, K., Rong, X., and Huang, Y. (2016). Widespread interspecies homologous recombination reveals reticulate evolution within the genus *Streptomyces*. *Mol. Phylogenet. Evol.* 102, 246–254. doi: 10.1016/j.ympev.2016.06.004
- Contreras-Moreira, B., and Vinuesa, P. (2013). GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13
- Corander, J., Waldmann, P., Marttinen, P., and Sillanpää, M. J. (2004). BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* 20, 2363–2369. doi: 10.1093/bioinformatics/bth250
- Csürös, M. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26, 1910–1912. doi: 10.1093/bioinformatics/btq315
- Doroghazi, J. R., and Buckley, D. H. (2010). Widespread homologous recombination within and between *Streptomyces* species. *ISME J.* 4, 1136–1143. doi: 10.1038/ismej.2010.45
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995
- Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2020). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 2020:1018. doi: 10.1093/nar/gkaa1018
- Girard, G., Willemse, J., Zhu, H., Claessen, D., Bukarasam, K., Goodfellow, M., et al. (2014). Analysis of novel *kitasatosporae* reveals significant evolutionary changes in conserved developmental genes between *Kitasatospora* and *Streptomyces*. *Antonie Van Leeuwenhoek* 106, 365–380. doi: 10.1007/s10482-014-0209-1
- Gonzalez, O., and Zimmer, R. (2008). Assigning functional linkages to proteins using phylogenetic profiles and continuous phenotypes. *Bioinformatics* 24, 1257–1263. doi: 10.1093/bioinformatics/btn106

- Goodfellow, M., Kämpfer, P., Busse, H.-J., Trujillo, M. E., Suzuki, K.-I., Ludwig, W., et al. (2012). *Bergey's Manual® of Systematic Bacteriology: Volume Five the Actinobacteria, Part A*. New York, NY: Springer.
- Hao, X., Yu, J., Wang, Y., Connolly, J. A., Liu, Y., Zhang, Y., et al. (2020). Zelkovamycins B-E, cyclic octapeptides containing rare amino acid residues from an endophytic *Kitasatospora* sp. *Org. Lett.* 22, 9346–9350. doi: 10.1021/acs.orglett.0c03565
- Hinger, I., Ansoorge, R., Musmann, M., and Romano, S. (2020). Phylogenomic analyses of members of the widespread marine heterotrophic genus *Pseudovibrio* suggest distinct evolutionary trajectories and a novel genus, *Polycladidibacter* gen. nov. *Appl. Environ. Microbiol.* 86:19. doi: 10.1128/AEM.02395-19
- Hopwood, D. A. (2007). *Streptomyces in Nature and Medicine: The Antibiotic Makers*. Oxford: Oxford University Press.
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mole. Biol. Evol.* 23, 254–267.
- Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9
- Kämpfer, P. (2012). Genus I. *Streptomyces* Waksman and Henrici 1943, 339<sup>AL</sup> emend. Witt and Stackebrandt 1990, 370 emend. Wellington, Stackebrandt, Sanders, Wolstrup and Jorgensen 1992, 159. *Actinobacteria* 5, 1455–1767.
- Kim, M. J., Roh, S. G., Kim, M. K., Park, C., Kim, S., and Kim, S. B. (2020). *Kitasatospora acidiphila* sp. nov., isolated from pine grove soil, exhibiting antimicrobial potential. *Int. J. Syst. Evol. Microbiol.* 70, 5567–5575. doi: 10.1099/ijsem.0.004450
- Kim, S. B., Lonsdale, J., Seong, C. N., and Goodfellow, M. (2003). *Streptacidiphilus* gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family *Streptomycetaceae* (Waksman and Henrici (1943)AL) emend. Rainey et al. 1997. *Antonie Van Leeuwenhoek* 83, 107–116. doi: 10.1023/a:1023397724023
- Labeda, D. P., Dunlap, C. A., Rong, X., Huang, Y., Doroghazi, J. R., Ju, K. S., et al. (2017). Phylogenetic relationships in the family *Streptomycetaceae* using multi-locus sequence analysis. *Antonie Van Leeuwenhoek* 110, 563–583. doi: 10.1007/s10482-016-0824-0
- Lee, M. D. (2019). GTToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164. doi: 10.1093/bioinformatics/btz188
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290
- Li, L., Stoekert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Li, Y., Sun, Z. Z., Rong, J. C., and Xie, B. B. (2021). Comparative genomics reveals broad genetic diversity, extensive recombination and nascent ecological adaptation in *Micrococcus luteus*. *BMC Genomics* 22:124. doi: 10.1186/s12864-021-07432-5
- Lin, M., and Kussell, E. (2019). Inferring bacterial recombination rates from large-scale sequencing datasets. *Nat. Methods* 16, 199–204. doi: 10.1038/s41592-018-0293-7
- Mandryk-Litvinkovich, M. N., Muratova, A. A., Nosonova, T. L., Evdokimova, O. V., Valentovich, L. N., Titok, M. A., et al. (2017). Molecular genetic analysis of determinants defining synthesis of 2,4-diacetylphloroglucinol by *Pseudomonas brassicacearum* BIM B-446 bacteria. *Appl. Biochem. Microb.* 53, 31–39. doi: 10.1134/S0003683817010124
- McDonald, B. R., and Currie, C. R. (2017). Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *MBio* 8:3. doi: 10.1128/mBio.00644-17
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11, 625–631. doi: 10.1038/nchembio.1890
- Mostowy, R., Croucher, N. J., Andam, C. P., Corander, J., Hanage, W. P., and Marttinen, P. (2017). Efficient inference of recent and ancestral recombination within bacterial populations. *Mol. Biol. Evol.* 34, 1167–1182. doi: 10.1093/molbev/msx066
- Navarro-Muñoz, J. C., Selem-Mojica, N., Mallowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* 16, 60–68. doi: 10.1038/s41589-019-0400-9
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nouioui, I., Klenk, H. P., Igual, J. M., Gulvik, C. A., Lasker, B. A., and McQuiston, J. R. (2019). *Streptacidiphilusbronchialis* sp. nov., a ciprofloxacin-resistant bacterium from a human clinical specimen; reclassification of *Streptomyces griseoplanus* as *Streptacidiphilus griseoplanus* comb. nov. and emended description of the genus *Streptacidiphilus*. *Int. J. Syst. Evol. Microbiol.* 69, 1047–1056. doi: 10.1099/ijsem.0.003267
- Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. doi: 10.1038/ismej.2017.126
- Ômura, S., Otoguro, K., Nishikiori, T., Oiwa, R., and Iwai, Y. (1981). Setamycin, a new antibiotic. *J. Antibiot.* 34, 1253–1256. doi: 10.7164/antibiotics.34.1253
- Ômura, S., Takahashi, Y., Iwai, Y., and Tanaka, H. (1982). *Kitasatospora*, a new genus of the order *Actinomycetales*. *J. Antibiot.* 35, 1013–1019. doi: 10.7164/antibiotics.35.1013
- Park, C. J., and Andam, C. P. (2020). Distinct but intertwined evolutionary histories of multiple *Salmonella enterica* Subspecies. *mSystems* 5:19. doi: 10.1128/mSystems.00515-19
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Parte, A. C. (2018). LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* 68, 1825–1829. doi: 10.1099/ijsem.0.002786
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sigui, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Suzuki, R., and Shimodaira, H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542.
- Takahashi, Y. (2017). Genus *Kitasatospora*, taxonomic features and diversity of secondary metabolites. *J. Antibiot.* 70, 506–513. doi: 10.1038/ja.2017.8
- Tidjani, A. R., Lorenzi, J. N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., et al. (2019). Massive gene flux drives genome diversity between sympatric *Streptomyces* conspecifics. *mBio* 10:19. doi: 10.1128/mBio.01533-19
- Vinuesa, P., Ochoa-Sánchez, L. E., and Contreras-Moreira, B. (2018). GET\_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Front. Microbiol.* 9:771. doi: 10.3389/fmicb.2018.00771
- Wellington, E. M., Stackebrandt, E., Sanders, D., Wolstrup, J., and Jorgensen, N. O. (1992). Taxonomic status of *Kitasatospora*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces* Waksman and Henrici 1943, 339<sup>AL</sup>. *Int. J. Syst. Bacteriol.* 42, 156–160. doi: 10.1099/00207713-42-1-156
- Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617.
- Yun, B. R., Malik, A., and Kim, S. B. (2020). Genome based characterization of *Kitasatospora* sp. MMS16-BH015, a multiple heavy metal resistant soil actinobacterium with high antimicrobial potential. *Gene* 733:144379. doi: 10.1016/j.gene.2020.144379

- Zhang, G., Gao, B., Adeolu, M., Khadka, B., and Gupta, R. S. (2016). Phylogenomic analyses and comparative studies on genomes of the *Bifidobacteriales*: identification of molecular signatures specific for the order *Bifidobacteriales* and its different subclades. *Front. Microbiol.* 7:978. doi: 10.3389/fmicb.2016.00978
- Zhang, Z., Wang, Y., and Ruan, J. (1997). A proposal to revive the genus *Kitasatospora* (Ōmura, Takahashi, Iwai, and Tanaka 1982). *Int. J. Syst. Bacteriol.* 47, 1048–1054. doi: 10.1099/00207713-47-4-1048
- Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., et al. (2014). PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* 30, 1297–1299. doi: 10.1093/bioinformatics/btu017

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Wang, Sun and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.